



Zweifache lineare Regression

1

Worum geht es in diesem Modul?

- Beispiel: Intelligenz, Bleibelastung und beruflicher Status
- Multiple lineare Regression mit zwei Regressoren
- Die bedingten Regressionen
- Eigenschaften des Residuums
- Identifikation der Regressionskoeffizienten
- Dichotome Regressoren
- Einfache und zweifache Regression und multiple lineare Regression mit zwei Regressoren
- Lineare Quasi-Regression



Beispiel: Intelligenz, Bleibelastung und beruflicher Status

2

In diesem Einführungsbeispiel befassen wir uns mit der Frage, ob eine erhöhte Bleibelastung der Umwelt zu einer Verminderung der Intelligenzleistungen bei Kindern führt.

In der Stichprobe fand sich zwischen dem *Logarithmierten Bleigehalt* (X) und dem *Verbalen Intelligenzquotienten* (Y) eine negative Korrelation von -0.14 . Wenn wir davon ausgehen, dass die Abhängigkeit des Regressanden Y vom Regressor X *linear* regressiv ist, dann lässt sich diese Abhängigkeit durch die Gleichung

$$E(Y|X) = \alpha_0 + \alpha_1 X$$

beschreiben, wobei $\alpha_1 < 0$. Der Verbale IQ ist dann also negativ linear regressiv abhängig vom Bleigehalt der Zähne.



Beispiel: Intelligenz, Bleibelastung und beruflicher Status

3

Neben den beiden Variablen X und Y wurde aber auch der *Berufliche Status der Eltern* der Kinder erfragt, der mit Z bezeichnet sei. Es stellt sich nun die Frage, ob bei gegebenem Wert z von Z (also bei festem beruflichen Status) noch eine lineare regressive Abhängigkeit der Variablen Y von X besteht. Die Daten der Stolberg-Studie legen nun nahe, dass Y von X bezüglich Z partiell linear regressiv *unabhängig* ist, d. h., dass die Gleichung

$$E(Y | X, Z) = \beta_0 + \beta_1 X + \beta_2 Z$$

gilt, wobei der Koeffizient

$$\beta_1 = 0.$$



Zweifache lineare Regression: Definition

4

Definition 9.1

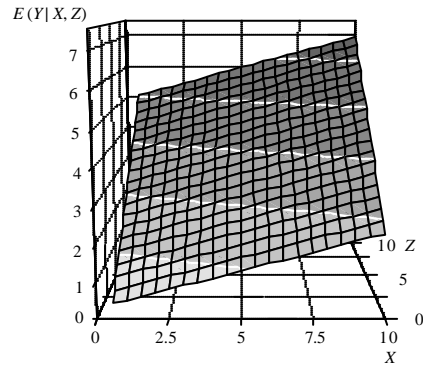
Seien X , Y und Z jeweils eindimensionale numerische Zufallsvariablen mit endlichen Erwartungswerten und Varianzen auf dem gleichen Wahrscheinlichkeitsraum. Dann heißt die Regression $E(Y | X, Z)$ *linear in* (X, Z) , wenn gilt:

$$E(Y | X, Z) = \beta_0 + \beta_1 X + \beta_2 Z, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R}.$$



Zweifache lineare Regression : Abbildung I

5



$$E_{Z=z}(Y | X) = (\beta_0 + \beta_2 z) + \beta_1 X,$$

Abbildung 9.1. Regressionsebene der Regression $E(Y|X, Z) = 0.3 + 0.2 \cdot X + 0.5 \cdot Z$. Jeder Punkt auf der Ebene ist (bei kontinuierlichen Regressoren X und Z) ein Wert $E(Y|X = x, Z = z)$ der Regression $E(Y|X, Z)$.



Die bedingten Regressionen

6

$$E_{Z=z}(Y | X) = (\beta_0 + \beta_2 z) + \beta_1 X,$$



Zweifache lineare Regression: Abbildung II

7

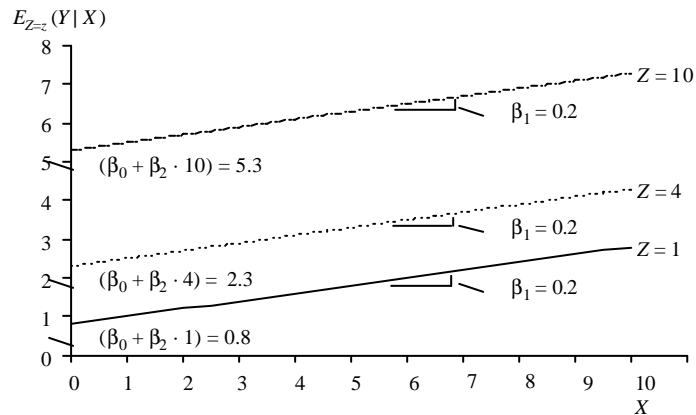


Abbildung 9.2. Die bedingten Regressionsgeraden bei bzgl. Z partieller linearer regressiver Abhängigkeit des Regressanden Y vom Regressor X . Dabei wird wie in Abbildung 9.1 die Regression $E(Y|X, Z) = 0.3 + 0.2 \cdot X + 0.5 \cdot Z$ zugrunde gelegt.



Determinationskoeffizient

8

$$R_{Y|X, Z}^2 - R_{Y|Z}^2$$

$$R_{Y|X, Z}^2 = \frac{\beta_1^2 \text{Var}(X) + \beta_2^2 \text{Var}(Z) + 2\beta_1\beta_2 \text{Cov}(X, Z)}{\text{Var}(Y)}$$



Das Residuum und seine Eigenschaften

9

Für das Residuum

$$\mathbf{e} := Y - E(Y | X, Z)$$

gelten, neben den bereit behandelten allgemeinen Eigenschaften, wie z.B.

$$E(\mathbf{e} | X, Z) = 0 \text{ und } E(\mathbf{e}) = 0,$$

insbesondere

$$E(\mathbf{e} | X) = E(\mathbf{e} | Z) = 0,$$

und

$$\text{Cov}(\mathbf{e}, X) = \text{Cov}(\mathbf{e}, Z) = 0.$$



Identifikation

10

$$\beta_0 = E(Y) - \beta_1 E(X) - \beta_2 E(Z)$$

$$\begin{aligned} \beta_1 &= \frac{\text{Var}(Z) \text{Cov}(X, Y) - \text{Cov}(X, Z) \text{Cov}(Y, Z)}{\text{Var}(X) \text{Var}(Z) - \text{Cov}(X, Z)^2} \\ &= \frac{\text{Std}(Y)}{\text{Std}(X)} \cdot \frac{\text{Kor}(X, Y) - \text{Kor}(X, Z) \text{Kor}(Y, Z)}{1 - \text{Kor}(X, Z)^2} \end{aligned}$$

$$\begin{aligned} \beta_2 &= \frac{\text{Var}(X) \text{Cov}(Z, Y) - \text{Cov}(X, Z) \text{Cov}(Y, X)}{\text{Var}(X) \text{Var}(Z) - \text{Cov}(X, Z)^2} \\ &= \frac{\text{Std}(Y)}{\text{Std}(Z)} \cdot \frac{\text{Kor}(Z, Y) - \text{Kor}(X, Z) \text{Kor}(X, Y)}{1 - \text{Kor}(X, Z)^2} \end{aligned}$$



Identifikationen: Dichotome Regressoren

11

Im speziellen Fall, in dem X und Z dichotom (zweiwertig) sind, lassen sich die Koeffizienten β_0 , β_1 und β_2 noch einfacher berechnen. In diesem Fall kann man die vier Gleichungen

$$E(Y | X = 1, Z = 1) = \beta_0 + \beta_1 + \beta_2$$

$$E(Y | X = 1, Z = 0) = \beta_0 + \beta_1$$

$$E(Y | X = 0, Z = 1) = \beta_0 + \beta_2$$

$$E(Y | X = 0, Z = 0) = \beta_0$$

für die vier bedingten Erwartungswerte $E(Y | X = x, Z = z)$ ableiten, und diese dann nach den unbekannten Koeffizienten β_0 , β_1 und β_2 auflösen:

$$\beta_1 = E(Y | X = 1, Z = 0) - E(Y | X = 0, Z = 0)$$

und

$$\beta_2 = E(Y | X = 0, Z = 1) - E(Y | X = 0, Z = 0)$$



Einfache und zweifache Regression: I

12

Theorem 9.1. Unter den gleichen Voraussetzungen wie in Definition 9.1 gilt folgendes: Ist $E(Y | X, Z) = \beta_0 + \beta_1 X + \beta_2 Z$ erfüllt und ist $\beta_2 = 0$ oder ist Z von X regressiv unabhängig,

$$E(Z | X) = E(Z),$$

dann folgt

$$E(Y | X) = \alpha_0 + \alpha_1 X,$$

wobei $\alpha_0 = \beta_0 + \beta_2 E(Z)$ und $\alpha_1 = \beta_1$ und weiter

$$\text{Cov}(X, Z) = 0,$$

$$\text{Var}[E(Y | X, Z)] = \beta_1^2 \text{Var}(X) + \beta_2^2 \text{Var}(Z).$$



Einfache und zweifache Regression: II

13

Gilt außerdem noch

$$E(X|Z) = E(X),$$

dann folgen auch

$$\text{Var}[E(Y|X, Z)] = \text{Var}[E(Y|X)] + \text{Var}[E(Y|Z)],$$

$$R_{Y|X,Z}^2 = R_{Y|X}^2 + R_{Y|Z}^2,$$

$$R_{Y|X}^2 = \text{Kor}(X, Y)^2$$

$$R_{Y|Z}^2 = \text{Kor}(Y, Z)^2$$



Einfache und zweifache Regression: III

14

Kann man nicht voraussetzen, dass Z von X regressiv unabhängig ist, und gilt stattdessen

$$E(Z|X) = \gamma_0 + \gamma_1 X,$$

so folgt dennoch

$$E(Y|X) = \alpha_0 + \alpha_1 X,$$

wobei $\alpha_0 := \beta_0 + \beta_2 \gamma_0$ und $\alpha_1 := \beta_1 + \beta_2 \gamma_1$, also

$$E(Y|X) = (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) X.$$



Multiple lineare Quasi-Regression: I

15

Definition 9.2. Unter den gleichen Voraussetzungen wie in Definition 9.1 handelt es sich bei der *zweifachen linearen Quasi-Regression*, die wir mit $Q(Y|X, Z)$ bezeichnen, um diejenige Linearkombination $\beta_0 + \beta_1 X + \beta_2 Z$ von X und Z , die folgendes erfüllt:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \mathbf{n} ,$$

$$E(\mathbf{n}) = 0 ,$$

und

$$\text{Cov}(\mathbf{n}, X) = \text{Cov}(\mathbf{n}, Z) = 0.$$



Multiple lineare Quasi-Regression: II

16

Definition 9.3. Unter den gleichen Voraussetzungen wie in Definition 9.1 können wir $Q(Y|X, Z)$ auch als diejenige Linearkombination von X und Z definieren, welche die folgende Funktion der reellen Zahlen b_0 , b_1 und b_2 , das *Kleinst-Quadrat-Kriterium* minimiert:

$$LS(b_0, b_1, b_2) = E[[Y - (b_0 + b_1 X + b_2 Z)]^2].$$

Diejenigen Zahlen b_0 , b_1 und b_2 , für welche die Funktion $LS(b_0, b_1, b_2)$ ein Minimum annimmt, seien mit β_0 , β_1 und β_2 respektive, bezeichnet. Die *zweifache lineare Quasi-Regression* ist dann definiert durch:

$$Q(Y|X, Z) := \beta_0 + \beta_1 X + \beta_2 Z$$